HSE

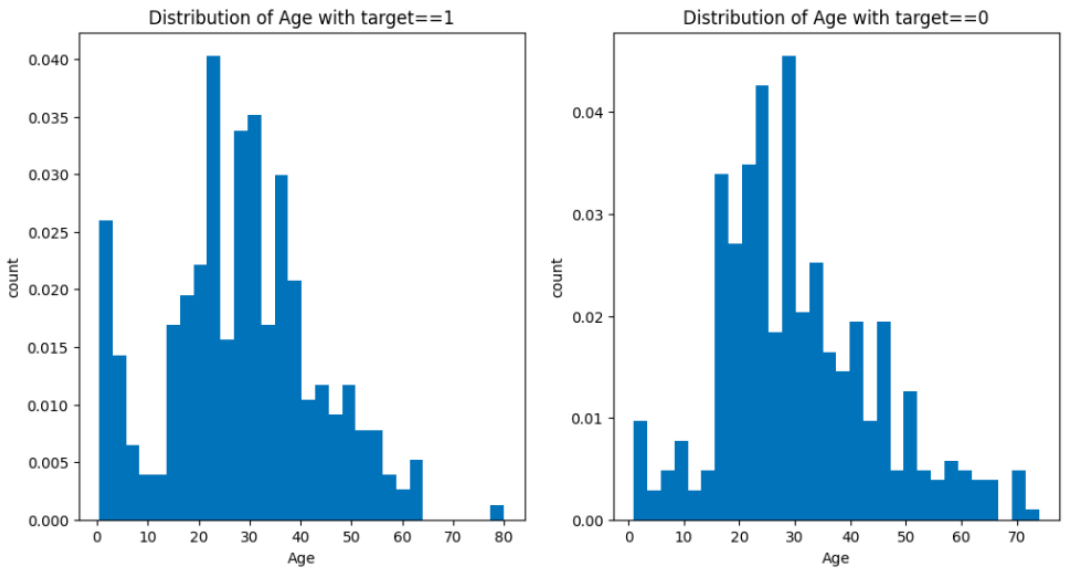
OSDA Lazy FCA project

Valerii Kropotin

December 10, 2024

I chose the Titanic dataset, which provides detailed information about passengers during the shipwreck. The binary classification is based on the fact whether the passenger survived the crash or not

1. Nominal scale features: Pclass, SibSp, Parch, Cabin, Embedded
2. Dichotomic scale features: Sex
3. Ordinal scale features: Age, Fare



1. Comparable results with KNN and decision free
2. Almost 50% of all samples has 0 positive and 0 negative classifiers
3. Random Forest and Logistic Regression performed best on the considered task

	True Positive	True Negative	False Positive	False Negative	Negative Predictive Value	False Positive Rate	False Discovery Rate	accuracy	precision	recall	f1
fca	41	98	12	28	0.777778	0.109091	0.405797	0.776536	0.773585	0.594203	0.672131
knn	46	94	16	23	0.803419	0.145455	0.333333	0.782123	0.741935	0.666667	0.702290
decision tree	47	93	17	22	0.808696	0.154545	0.318841	0.782123	0.734375	0.681159	0.706767
logistic regression	48	98	12	21	0.823529	0.109091	0.304348	0.815642	0.800000	0.695652	0.744186
random forest	50	97	13	19	0.836207	0.118182	0.275362	0.821229	0.793651	0.724638	0.757576
xgboost	45	97	13	24	0.801653	0.118182	0.347826	0.793296	0.775862	0.652174	0.708661